Gaze Guidance through Peripheral Stimuli

Andrei Lintu, Noëlle Carbonell

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Gaze Guidance through Peripheral Stimuli

Andrei Linţu — Noëlle Carbonell

N° ????

July 2009

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--????--FR+ENG

Gaze Guidance through Peripheral Stimuli

Andrei Linţu , Noëlle Carbonell

Thème COG — Systèmes cognitifs
Équipe-Projet Merlin

Rapport de recherche n° ???? — July 2009 — 14 pages

Abstract: Guiding gaze using near-peripheral vision improves gaze-contingent-display efficiency by reducing display response latency. We propose a new approach for controlling exploration of static displays through near-peripheral stimuli, and report results of an evaluation of its effectiveness. 10 participants viewed full screen displays of 60 blurred pictures (Gaussian filtering). As soon as a fixation (first strategy) or a gaze sample (second strategy) was detected next to the current stimulus, the area surrounding it was deblurred. An image was totally deblurred when all stimuli had thus attracted the users gaze.

Stimuli are blinking deblurred circles (radius: 1° visual angle). They appear in predefined positions on the screen, one at a time. For each picture, successive stimulus positions on the screen reproduce observed gaze patterns. The current stimulus is visible only if the visual angle between its position on the screen and the position of the users current fixation is superior to 8° (to avoid users noticing it) and inferior to 14° (near-periphery upper limit). Eye movements are detected through an ASL-H6 eye-tracker (120 Hz). Stimulus saliency is estimated, for each picture and stimulus, from contrast ratio and sum of squared differences between blurred and deblurred area around the stimulus.

Key-words: eye-tracking, peripheral vision, gaze guiding, vision

Centre de recherche INRIA Nancy – Grand Est
LORIA, Technopôle de Nancy-Brabois, Campus scientifique,
615, rue du Jardin Botanique, BP 101, 54602 Villers-Lès-Nancy (France)
Téléphone : +33 3 83 59 30 00 — Télécopie : +33 3 83 27 83 19

Gaze Guidance through Peripheral Stimuli

Résumé : Ce document . . . de l'equipe Merlin a l'INRIA Nancy Grand Est.

Mots-clés : eye-tracking, peripheral vision, gaze guiding, vision

1 Introduction

The gaze guiding framework described in this technical report was presented at the 15th European Conference on Eye Movements, 24th August 2009, Southampton, UK, as an oral presentation with the title: "Gaze guidance through peripheral stimuli"[1].

Gaze guiding can be used in a wide range of applications. Fields of use range from: education, transfer of knowledge from expert to novice, better understanding of art, advertising or numerous applications for driving a motor vehicle, just to name a few. The before mentioned applications can be implemented in the fields of augmented or virtual reality, as well as for attentive displays. We present here a gaze guiding application which can be used in the field of gaze-contingent displays, which are a subcategory of attentive displays. For a general overview paper on attentive displays the reader should refer to [2].

A general problem of gaze-contingent displays is their limited response latency. We propose a gaze guiding method using near-peripheral stimuli which can decrease this latency by correctly anticipating the users gaze patterns. We also propose a novel and simple way to compute the saliency of the stimuli used in our framework.

Gaze contingent displays can be used in conjunction with the visualization of large data sets on large display devices in cases where interactivity is important. Even with today's powerful graphics hardware, very large data sets viewed at high resolution can cause problems at full screen display which makes a gaze contingent solution feasible in the advent of today's high quality and precise, non-intrusive eye trackers.

We will give at this point short descriptions of concepts from vision science which are important in the design and understanding of our proposed framework. *Bottom-up* processing, or data-driven processing refers to the type of information processing in the human visual system which can be described as a flowchart where the retinal image stays at the bottom and its subsequent, more processed interpretations are at higher levels. Conversely, *top-down* processing or hypothesis driven processing refers to processes that work in the other direction, with higher level representations as input and producing lower and lower level representations as output. It is an accepted fact in the pshycophisics literature that most processes in the early stages of visual processing are bottom-up, while there are many top-down processes present at a later level of visual processing - mainly where people have prior existing models of the viewed scene [3, pp. 84–85].

Another interesting subdivision of the human visual system is the classification in the *where* and *what* system. The where system consists of larger cell subdivision situated mostly in the periphery on the retina and is responsible for motion perception, depth perception, spatial organization and figure/ground segregation. The where system is color-blind, fast, has low acuity and high contrast sensitivity. The what system has a small-cell subdivision and is mostly placed in the foveal region of the retina. It is responsible for object and face recognition and color perception. The what system is color selective, slow, has high acuity and low contrast sensitivity. For more details on the where and what system, please refer to [4, Chapter 4].

The main goal of our framework is to guide gaze with stimuli which are not consciously perceived by the subject. We achieve this goal by showing high

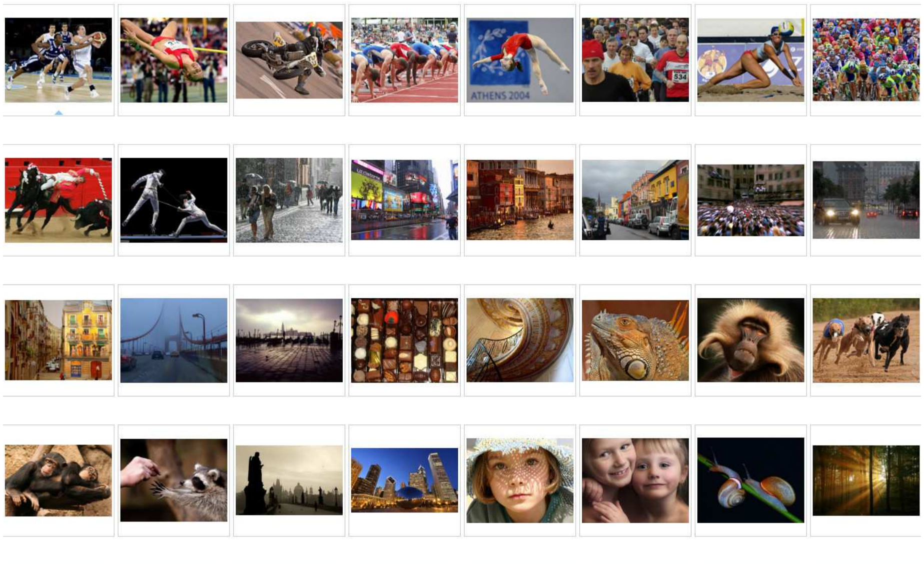

Figure 1: The first 30 images in our collection

resolution patches on a low resolution image viewed by the subject, placed in the peripheral region of the visual field, i.e. we want to stimulate the *where* system. The user views a set of images and its task is to deblurr all images, successively. If the users gaze is attracted by the current stimulus, an area associated with the stimulus is deblurred and the next stimulus is turned on.

We propose two different strategies for guiding: *with* (see Figure 5) and *without* anticipation (see Figure 6). The strategy with anticipation is based on the idea that most of the subjects saccades are preprogrammed and that we can already deblurr the area associated with the stimulus if the direction of the saccade points towards it. In the case of the strategy without anticipation, we only deblurr the area corresponding to the current stimulus when an eye tracker sample is already close to the stimulus.

The position of most our stimuli is given by the position of fixations from a previous pilot study consisting of a free exploration of selected 60 images. A thumbnail overview of our image collection is displayed in Figures 1 and 2. An example of a recorded gaze pattern for one image is given in Figure 3. The white circles indicate fixations, with the circle radius proportional to fixation length. The goal of our experiment is to deblur the images in their totality, thus stimuli have to be present even on the periphery of the images. As it can be seen in Figure 3, gaze is mostly attracted to the central parts of the image and this is why we place additional stimuli in the peripheral region, following a spiral distribution.

As already mentioned earlier, each stimulus has its associated area surrounding it which gets deblurred, once the stimulus attracts the subject's gaze. For simplicity, we divide our input images in a grid of cells, and each cell has a stimulus associated with it, as shown in Figure 4. Each cell is labeled with the same color, as well as number of its corresponding stimulus.

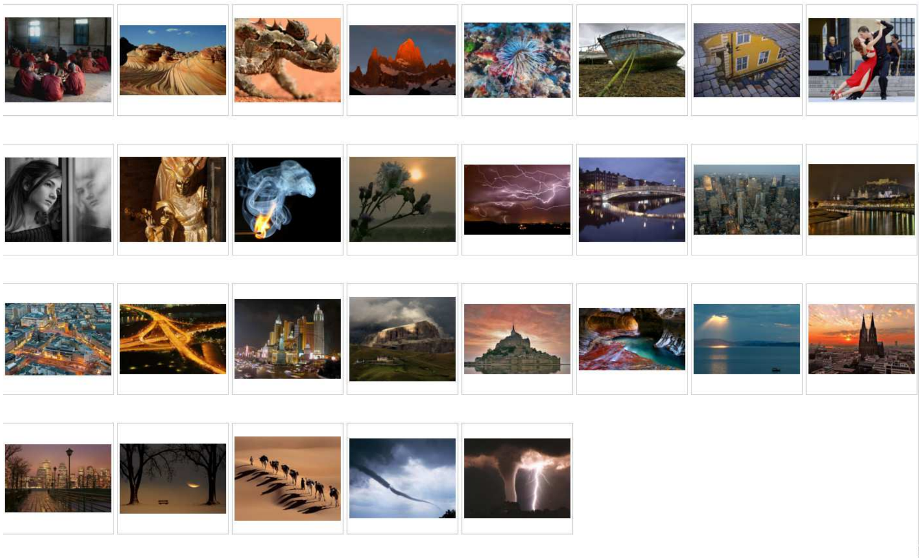

Figure 2: The next 30 images in our collection

To assess the saliency of our stimuli we propose a novel model to measure stimulus saliency which uses contrast as one of its metrics. Mainly used for phsycophysics experiments, the Michelson contrast [5, pp. 31], is defined as:

$$C = \frac{I_{min} - I_{max}}{I_{min} + I_{max}} \tag{1}$$

It is mainly used in experiments measuring contrast sensitivity. A definition better adapted for contrast in complex images was proposed by Peli [6]. An improved version of the contrast in complex images was proposed by Mantiuk et al. [7]. We use in our stimulus saliency computations a simplified, yet concise version of contrast measurement by Rizzi et al. [8].

For the reader interested in the human visual system in general, good introductory books on the general topic of human visual perception are [4, 9]. For a neuroscience view on the human visual system, the reader should refer to [10]. Another solid source on visual perception is [3].

2 Related Work

Dorr et al. [11] examine a system to guide eye movements using two types of stimuli in the periphery. The subjects have to view short videos of natural scenes; the success and "non-detection" of stimuli for guidance is evaluated.

Parkhurst et al. [12] use a bottom-up model of visual processing based on the two stage processing framework proposed in the literature: *early vision* and *center surround*. The authors goal is to provide a framework which can predict fixation locations using a saliency map. They use a two stage multi-resolution approach two compute saliency maps and use them afterwards to evaluate the correlation between areas of high saliency and fixation location while the subjects viewed natural and artificial images. The authors conclude that bottom-up

Figure 3: Image showing scan-paths and fixations used for stimulus placement

mechanisms contribute significantly to attentional guidance under natural viewing conditions.

In [13] the authors investigate the possible correlation between the local contrast and fixation location for several types of natural images while free viewed by subjects. Analysis is made for image patches of different sizes, but the biggest difference between the contrast of fixated patches compared to random patches is attained at a patch size of 1 ° visual angle.

Contrast sensitivity in peripheral vision is scrutinized in [14]. The authors concentrate on two tasks: resolution and detection of stimuli. Several experiments are described and as an important general conclusion the authors state that the detection task is much less affected by the mechanisms that limit resolution in the peripheral vision.

Itti et al. [15] present a saliency-based model of visual attention. Their model includes center-surround operations, a multi resolution Gaussian pyramid as well as a model for the orientation of features in the image. The presented model reproduces well human performance for pop-out tasks as well as detects most regions of high saliency in the input images.

In the fields of computer graphics and image processing, an algorithm for the assessment of image fidelity was proposed by Daly [16], called the *Visible Differences Predictor*. The algorithm tries to asses at which degree any differences physically present in images due to image processing or displaying are also visible differences. It is a complex framework which can be used in many fields for

Figure 4: Our stimuli are placed at the numbered positions marked by the red circles. All corresponding surrounding cells are labeled with the same number

image quality assessment and encompasses several models of the human visual system.

3 Experimental Setup

We will describe here into more detail our experimental setup. The subjects observe the sequence of images which have only a single deblurred area in the center at the beginning. If the given conditions are satisfied, and the current stimulus is visible and also attracts the users gaze the corresponding area is deblurred and merged with the already deblurred regions. The goal for the user is to deblur the hole image; we have in mean around 10 stimuli per image (depending on the complexity of each given image). After each image is deblurred, the subject is presented a questionnaire regarding several aspects of the process (if the process was annoying, or if the user likes the image visually).

The current stimulus is visible only if the visual angle between its position on the screen and the position of the users current fixation is superior to 8° - which puts the stimulus already in the foveal or para-foveal vision and inferior to 14° which is the limit of the near-peripheral vision. Eye movements are detected through an ASL-H6 eye-tracker working at 120 Hz.

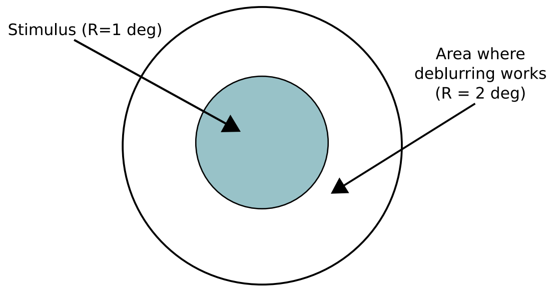

Figure 5: A schematic description of the mechanism used for our *without anticipation* strategy. The center shaded circle represents the area of 1 ° visual angle occupied by the stimulus. The outer circle, of 2 ° visual angle is the area in which if an eye-tracker sample falls the deblurring is activated

Our stimuli are blinking, high resolution patches of 1 ° radius overlayed on the low resolution image which are only present in the peripheral field of vision. The blinking frequency of our stimuli as a function of distance from the current fixation to the stimulus was adapted from [11] and [17].

To be able to compare our two proposed strategies we use in random way one of them for each image presented to the subject. This way we can avoid the subjects learning of a particular strategy and have more significant results.

We asses the effectiveness of our proposed guiding method by measuring as principal indicators: the time in which gaze is attracted to every stimulus, and the distance from the last fixation to the current stimulus.

All images where low pass filtered using a simple Gaussian Blurring kernel of 20 pixels in GIMP. The blurring was selected by trial and error in such a way as to keep in most of the cases the contrast between the stimuli and its surroundings as low as to be just *detected*, but not *resolved*, see [14].

3.1 Guidance

We propose two different guiding strategies. For the strategy *with anticipation*, a gaze sample has to fall further than half the distance between the previous fixation and the stimulus, as illustrated in Figure 6. By using this strategy we try to anticipate the subjects next fixation and additionally avoid the subjects noticing the current stimulus. In the case of the strategy *without anticipation*, Figure 5, a gaze sample has two fall into the circle of 2 ° visual angle radius surrounding the current stimulus. Because one of the goals of our framework is to keep stimuli as unperceived by subjects as possible in this case the stimulus will be invisible because it falls in the subjects foveal vision. Our system deblurs the area corresponding to the stimulus even if it is invisible, in the case the sample meets the strategy's constraints. We rely on the fact that saccades are ballistic movements and once begun, their trajectory cannot be altered [3, pp. 523]. If the peripheral stimulus was perceived by the subject, there is a good chance that the next fixation will land in the circle of 2 ° visual angle surrounding the current fixation.

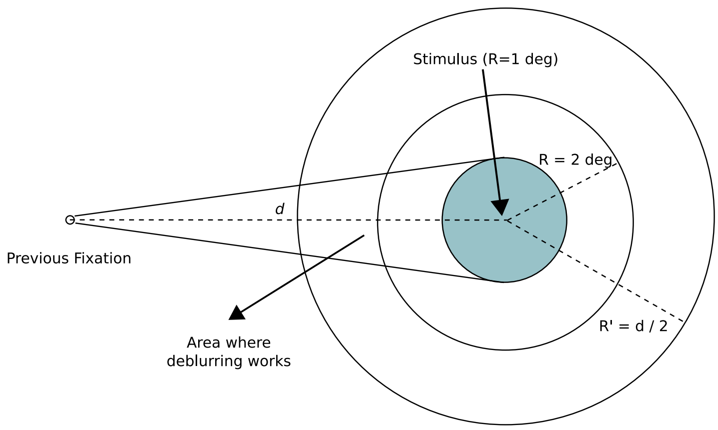

Figure 6: A schematic description of the mechanism used for our *with anticipation* strategy. The center shaded circle represents the area of 1° visual angle occupied by the stimulus. If the subjects gaze is attracted towards the stimulus from the previous fixation, deblurring occurs when an eye-tracker sample is at least at half the distance between the previous fixation and the stimulus

4 Stimulus Saliency

The stimuli used in our framework are circular patches with 1° visual angle radius. They are high resolution patches overlayed on the low resolution image, displayed at salient positions for each image, as described in Section 1.

We propose a novel measure for stimulus saliency based on two metrics: *contrast ratio* and *sum of squared differences* between the corresponding high resolution and low resolution patches for each stimulus position. This saliency model is much more simplified than previously proposed models such as the one by Parkhurst et al. [12] or Itti et al. [15], but has the advantage that it can be computed fast and it serves as an additional tool for the fast evaluation of the results of our effectuated experiments.

To measure the contrast ratio between 2 corresponding low and high resolution patches around all stimuli in the 60 images we use an GNU Octave [18] implementation of the contrast measurement method proposed by Rizzi et al. [8]. This method computes a local and multi-scale contrast for a given natural image by constructing a pyramid of sub-sampled images and computing the contrast at every level. The pyramid is built by dropping every second row and column from the input image at each level, for $N = [log_2 min(m, n)]$ levels, where $I(m, n)$ is the input image. After the input image I is de-constructed, the contrast for every pixel (x, y) is computed as the average difference between its value and that of the eight nearest neighbors. All these values are afterwards averaged to obtain an average contrast for every given resolution level. A global contrast can be computed by finally averaging the computed contrasts for every given level. The equation for computing the global contrast for the input image I

following all steps described earlier is as follows:

$$C(I) = \sum_{\forall\ level} \left(\frac{\sum_{\forall\ pixel} \left(\frac{\sum_{8-neigh} \frac{|P_i - P_j|}{8}}{number\ of\ pixels} \right)}{number\ of\ levels} \right) \tag{2}$$

Using the contrast measurement method described above, we systematically compute the global contrast for corresponding patches from the high resolution and blurred images from the regions surrounding all stimuli. Having the global contrast for every high resolution as well as blurred patch, we then compute the contrast ratio between each corresponding patch. For each image in our collection we plot the contrast ratios, to be able to asses stimulus saliency based on our contrast metric, as presented in Figure 7. The corresponding image for the computed saliencies is shown in Figure 8. For most cases the proposed contrast ratio measure can be used as a good indicator of stimulus saliency, however there are exceptions where an additional measure of stimulus strength has to be computed in order to be sure about the saliency of the stimulus.

We propose to use the *sum of squared differences* as a measure for additional investigation of stimulus saliency. For every image patch I around the predefined stimuli for each image, we compute the sum of squared differences (SSD) as:

$$ssd(I) = \sum_{\forall\ pixel} \sqrt{(I_{blurred}(m,n) - I_{high\ resolution}(m,n))^2} \tag{3}$$

where $I_{blurred}$ and $I_{high\ resolution}$ are the two corresponding 150 x 150 pixel regions around each stimulus from the blurred and high resolution image. An example of this metric for Figure 8 is also presented in Figure 7, alongside the contrast ratio metric.

5 Conclusion and Future Work

We presented a framework which demonstrates efficient gaze guiding using peripheral stimuli. Our system consists of a personal computer system running the custom software for displaying and free-viewing of natural images by subjects and the interface with an *ASL-H6* eye-tracker working at 120 Hz. The images are first displayed in a blurred, i.e. low pass filtered version with a single high resolution patch of 3° visual angle in the middle and the user deblurs the image step by step by following the stimuli in the peripheral vision with its gaze. The deblurring is done using one of the two proposed strategies for gaze guiding: *with* and *without anticipation*. The images viewed by the subjects are divided in a grid and each stimulus is assigned a region including the grid cell containing the stimulus plus other surrounding grid cells. Once the subjects gaze is attracted by a stimulus, its corresponding region is deblurred, i.e. the region is displayed in high resolution. The deblurring process ends for every image once the whole image is displayed in high resolution. The two guiding strategies proposed by us differ mainly by the conditions by which deblurring is effectuated.

We also presented a novel method to determine stimulus saliency which computes in parallel the contrast ratios and the sum of squared differences

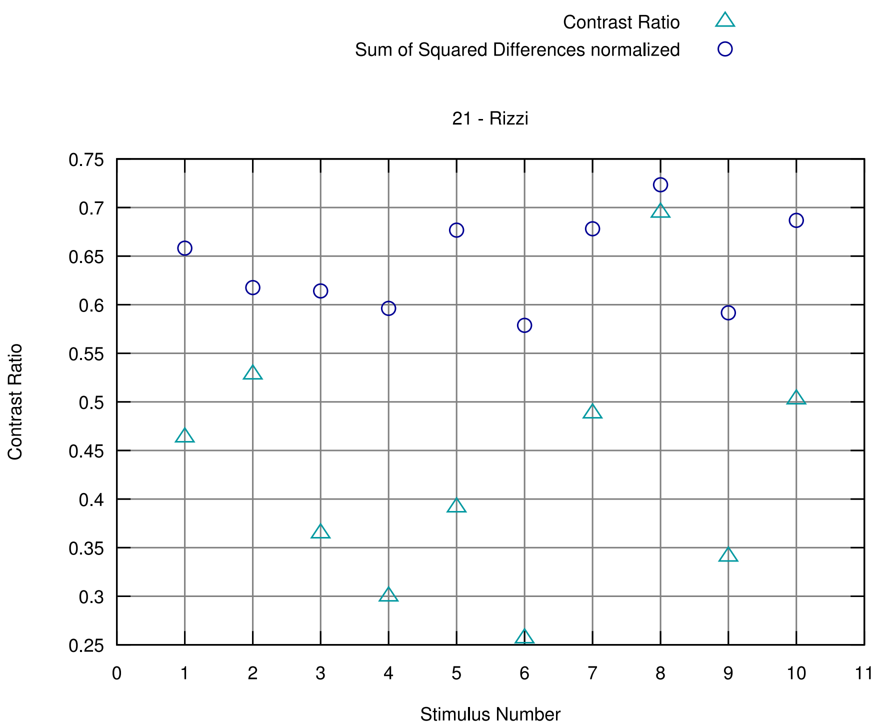

Figure 7: Stimulus saliency measures for a selected image. We measure the *contrast ratio* and *sum of squared differences* between corresponding patches

between corresponding stimuli in a high resolution, respectively a blurred image. The proposed measure can be used to asses stimulus saliency also for other similar applications.

As future direction we could include in the computation of stimulus saliency the eccentricity of stimuli or the local lightness adaptation level of the subject for every given stimulus. In the current implementation of our framework, all blurred images are precomputed with a given Gaussian kernel. Thus, the stimuli, which are the high resolution patches appearing on the blurred images have all an equal low pass filter applied. A possible improved system could take into account local adaptation levels of the subjects visual system and apply a locally adaptive blurring such that the stimuli stay at the detection threshold.

6 Acknowledgments

I would first like to thank Prof. Noëlle Carbonell who gave me the opportunity to work on this project during my post-doc stay here at LORIA. Sadly, Noëlle passed away in April 2009; she was a wonderful person and a dedicated scientist. I will miss her very much; I enjoyed a lot the lively scientific as well as non-scientific discussions with her.

A lot of our thanks go to Daniel Gepner, who worked earlier on this topic and performed the pilot studies to acquire subject scan paths while free-viewing the

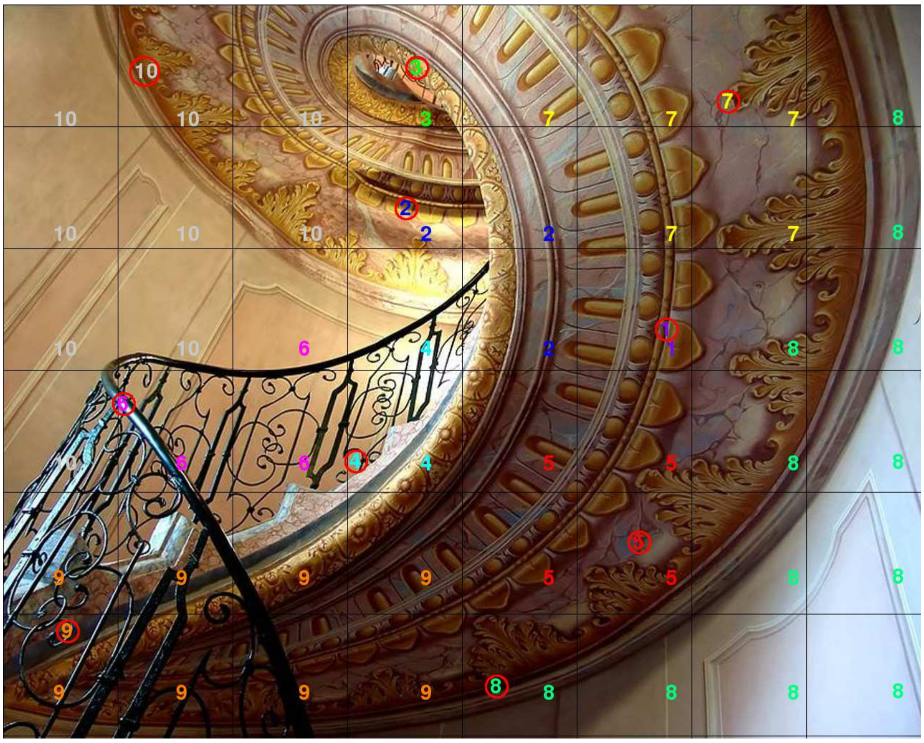

Figure 8: Image with the position of stimuli highlighted corresponding to the stimulus saliency measured in Figure 7

images included in our framework. I would also like to thank Olivier Christmann and Jerome Simonin for being such nice and helpful colleagues and good friends during my time at Loria. Last but not least I would like to thank my dear friend Cristian Popi, for all the interesting and inspiring discussions we had in Nancy.

Contents

References

[1] A. Linţu and N. Carbonell, "Gaze guidance through peripheral stimuli," in *15th European Conference on Eye Movements*, p. 37, 2009. Oral presentation.

[2] P. Baudisch, D. DeCarlo, A. Duchowski, and W. Geisler, "Focusing on the essential: considering attention in display design," *Communications of the ACM*, vol. 46, no. 3, pp. 60–66, 2003.

[3] S. Palmer, *Vision science: photons to phenomenology.* MIT Press Cambridge, MA., 1999.

[4] M. Livingstone, *Vision and Art: The Biology of Seeing.* Harry N. Abrams, 2002.

[5] A. Michelson, *Studies in Optics.* Dover Publications, 1995.

[6] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[7] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Transactions on Applied Perception (TAP)*, vol. 3, no. 3, pp. 286–308, 2006.

[8] A. Rizzi, T. Algeri, G. Medeghini, and D. Marini, "A proposal for contrast measure in digital images," in *CGIV 2004–Second European Conference on Color in Graphics, Imaging and Vision*, 2004.

[9] D. Marr, *Vision.* W. H. Freeman and Company, 1982.

[10] D. Hubel, *Eye, brain, and vision.* Scientific American Library New York, 1988.

[11] M. Dorr, T. Martinetz, K. Gegenfurtner, and E. Barth, "Guidance of Eye Movements on a Gaze-Contingent Display," *Dynamic Perception Workshop of the GI Section" Computer Vision", UJ Illg, HH Bülthoff, and HA Mallot, Eds*, 2004.

[12] D. J. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, 2002.

[13] D. J. Parkhurst and E. Niebur, "Scene content selected by active vision," *Spatial Vision*, vol. 16, no. 2, pp. 125–154, 2003.

[14] L. N. Thibos, D. Still, and A. Bradley, "Characterization of spatial aliasing and contrast sensitivity in peripheral vision," *Vision Research*, vol. 36, no. 2, pp. 249–258, 1996.

[15] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pp. 1254–1259, 1998.

[16] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Proceedings of SPIE*, vol. 1666, p. 2, SPIE, 1992.

[17] M. Dorr, T. Martinetz, M. Böhme, and E. Barth, "Visibility of temporal blur on a gaze-contingent display," *Proceedings of the 2nd symposium on Appied perception in graphics and visualization*, pp. 33–36, 2005.

[18] J. Eaton, "GNU Octave." available from `http://www.gnu.org/software/octave/`.

Centre de recherche INRIA Nancy – Grand Est
LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Centre de recherche INRIA Futurs : Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399